

Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics

Vuong Quoc Hoang Ngo, François Enault, Cédric Midoux, Mahendra Mariadassou, Olivier Chapleur, Laurent Mazéas, Valentin Loux, Théodore Bouchez, Mart Krupovic, Ariane Bize

▶ **To cite this version:**

Vuong Quoc Hoang Ngo, François Enault, Cédric Midoux, Mahendra Mariadassou, Olivier Chapleur, et al.. Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics. Environmental Microbiology, 2022, 10.1111/1462-2920.16120 . hal-03727436

HAL Id: hal-03727436
https://hal.inrae.fr/hal-03727436

Submitted on 19 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

Received: 8 April 2022 | Revised: 1 June 2022 | Accepted: 29 June 2022

DOI: 10.1111/1462-2920.16120

RESEARCH ARTICLE

environmental microbiology sfam

Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics

Vuong Quoc Hoang Ngo[1] | **François Enault**[2] | **Cédric Midoux**[1,3,4] | **Mahendra Mariadassou**[3,4] | **Olivier Chapleur**[1] | **Laurent Mazéas**[1] | **Valentin Loux**[3,4] | **Théodore Bouchez**[1] | **Mart Krupovic**[5] | **Ariane Bize**[1]

[1]Université Paris-Saclay, INRAE, PRocédés biOtechnologiques au Service de l'Environnement, Antony, France

[2]Université Clermont Auvergne, CNRS, LMGE, Clermont-Ferrand, France

[3]Université Paris-Saclay, INRAE, MaIAGE, Jouy-en-Josas, France

[4]Université Paris-Saclay, INRAE, BioinfOmics, MIGALE Bioinformatics Facility, Jouy-en-Josas, France

[5]Institut Pasteur, Université de Paris, CNRS UMR6047, Archaeal Virology Unit, Paris, France

Correspondence
Ariane Bize, Université Paris-Saclay, INRAE, PRocédés biOtechnologiques au Service de l'Environnement, 92761 Antony, France.
Email: ariane.bize@inrae.fr

Funding information
Agence Nationale de la Recherche, Grant/Award Numbers: ANR-17-CE05-0011, ANR-20-CE20-0009

Abstract

Diversity of viruses infecting non-extremophilic archaea has been grossly understudied. This is particularly the case for viruses infecting methanogenic archaea, key players in the global carbon biogeochemical cycle. Only a dozen of methanogenic archaeal viruses have been isolated so far. In the present study, we implemented an original coupling between stable isotope probing and complementary shotgun metagenomic analyses to identify viruses of methanogens involved in the bioconversion of formate, which was used as the sole carbon source in batch anaerobic digestion microcosms. Under our experimental conditions, the microcosms were dominated by methanogens belonging to the order Methanobacteriales (*Methanobacterium* and *Methanobrevibacter* genera). Metagenomic analyses yielded several previously uncharacterized viral genomes, including a complete genome of a head-tailed virus (class *Caudoviricetes*, proposed family *Sperloviridae*, *Methanobacterium* host) and several near-complete genomes of spindle-shaped viruses. The two groups of viruses are predicted to infect methanogens of the *Methanobacterium* and *Methanosarcina* genera and represent two new virus families. The metagenomics results are in good agreement with the electron microscopy observations, which revealed the dominance of head-tailed virus-like particles and the presence of spindle-shaped particles. The present study significantly expands the knowledge on the viral diversity of viruses of methanogens.

INTRODUCTION

Viruses infecting archaea represent one of the most unique parts of the global virosphere (Krupovic et al., 2018). Despite the limited number of archaeal viruses described so far compared to bacterial viruses, archaeal viruses show a great diversity of gene content and morphological properties. In particular, several morphotypes are specific to archaeal viruses, showing no similarity to viruses infecting bacteria and eukaryotes, such as bottle-shaped (*Ampullaviridae*), coil-shaped (*Spiraviridae*), or spindle-shaped (*Bicaudaviridae*, *Fuselloviridae*, *Halspiviridae*, *Thaspiviridae*) ones. Archaeal viruses are currently classified into 33 families (Baquero et al., 2020; Liu et al., 2021), including cosmopolitan icosahedral viruses and archaea-specific viruses.

Methanogenic archaea play a major role in carbon cycling at the global scale, through methanogenesis.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

Known methanogenic archaea are currently grouped into eight different orders in the phyla Euryarchaeota and Halobacteriota, and a few candidatus taxa in the Euryarchaeota, Halobacteriota, and in the TACK group (Evans et al., 2019; Lyu et al., 2018) These were isolated from very diverse natural ecosystems such as wetlands, termite, human and livestock digestive tracts, rice fields, and deep-sea hydrothermal vents, and also from anaerobic digesters (ADs) (Lyu et al., 2018). Indeed, their unique metabolic features are exploited in AD processes (Ahring, 2003) for valorization of organic waste and effluents into methane-rich biogas, a renewable energy source. Among archaeal viruses, 10 have been reported to infect methanogenic archaea (methanogens) (Krupovič et al., 2010a; Meile et al., 1989; Molnár et al., 2020; Nölling et al., 1993; Pfister et al., 1998; Thiroux et al., 2021; Weidenbach et al., 2017; Weidenbach et al., 2021; Wolf et al., 2019; Wood et al., 1989). In addition, several proviruses integrated in the genomes of diverse methanogens have been described (Krupovič et al., 2010; Krupovič & Bamford, 2008). Almost all of the viruses of methanogens described so far have been isolated from AD samples (Table S1). A total of five head-tailed viruses or proviruses originate from thermophilic ADs, all infecting Methanobacteriales hosts: ΨM1 and ΨM2 (family *Leisingerviridae*; siphovirus morphology) (Liu et al., 2021; Meile et al., 1989; Pfister et al., 1998) and related defective provirus ΨM100 (Luo et al., 2001) infect *Methanothermobacter* strains, whereas ΦF1 (unclassified) and ΦF3 (unclassified; siphovirus morphology) (Nölling et al., 1993) infect *Methanobacterium* species. Moreover, four viruses or virus-like particles (VLPs) have been isolated from mesophilic ADs: *Methanobacterium*-infecting virus Drs3 (family *Anaerodiviridae*; siphovirus morphology) (Liu et al., 2021; Wolf et al., 2019), Blf4 (unclassified; siphovirus morphology), infecting *Methanoculleus* strains (Methanomicrobiales) (Weidenbach et al., 2021), MetSV (unclassified) (Weidenbach et al., 2017), a spherical virus infecting *Methanosarcina* strains (Methanosarcinales), and finally the oblate or spindle-shaped *Methanococcus* (Methanococcales)-infecting A3 VLPs (Wood et al., 1989). MFTV1, an unclassified temperate head-tailed virus (siphovirus morphology), has been induced from a *Methanocaldococcus fervens* AG86 strain (Methanococcales) isolated from deep-sea hydrothermal vents (Thiroux et al., 2021); it is the first characterized virus infecting hyperthermophilic methanogens. In addition, MetMV (unclassified) (Molnár et al., 2020) has been suggested to infect mesophilic *Methanosarcina* strains, but this still needs to be confirmed. For the other four known orders of methanogens (Evans et al., 2019), no viruses have been isolated so far. Thus, the diversity of viruses infecting methanogenic archaea remains largely unexplored.

Metagenomics can provide a less biased view on the diversity of viruses infecting methanogens, by circumventing the challenge of cultivating some of the methanogenic archaeal strains, as well as biases associated with virus isolation. In such context, AD ecosystems prove to be particularly well suited as they are relatively easy to access to, and to establish in laboratory reactors. Moreover, they encompass methanogenic archaea from several orders, such as Methanobacteriales, Methanomicrobiales and Methanosarcinales (Evans et al., 2019; Lin et al., 2016). Yet, identifying viruses infecting methanogens is challenging in AD metagenomes due to the complexity of the catalytic microbial communities, dominated by diverse bacteria, and due to usually low proportions of methanogenic archaea in AD processes.

In the present study, an original experimental approach was applied with the aim of favouring the enrichment of AD microbial communities in methanogens, to help the discovery of their viruses in metagenomes. To this end, AD microcosms were fed with ^{13}C-labelled formate, one of the known substrates for methanogenesis through the following equation (Sun et al., 2021):

$$4HCOO^- + H^+ + H_2O \rightarrow CH_4 + 3HCO_3^- \left(\Delta G^0 : -170.2 kJ/mol\right)$$

In addition to favouring methanogens, such an experimental approach has the advantage of preserving the AD process and a certain level of microbial diversity. In particular, it can enable to reach higher proportions of methanogens, and it is also compatible with the presence of several methanogenic species or genera, offering the possibility of a relatively broad view on the diversity of viruses of methanogens. Another benefit of this method is the possibility to identify DNA viruses targeting microorganisms that actively assimilated the substrate. Indeed, within complex microbial communities, not all of the microorganisms are active or involved in the degradation of a specific substrate. Thus, identifying active microorganisms, and their associated viruses, are key issues in linking viral diversity with functional aspects. Hence, we coupled stable isotope probing (SIP) (Radajewski et al., 2000), applied to the total cellular DNA, to *in silico* host prediction of the viral contigs. Host prediction was applied by using the cellular metagenomes obtained from the heavy (^{13}C-enriched) and light (not enriched in ^{13}C) DNA sequences, which were used to build host databases. Whenever the predicted host was identified in the heavy cellular DNA fraction, we considered it as evidence that the virus infected active microorganisms, assimilating the ^{13}C-formate.

This original coupling (SIP and bioinformatic analyses for host prediction) led to the discovery of several previously uncharacterized genomes of DNA viruses of methanogens. In particular, they included two contigs likely representing new families of spindle-shaped

viruses, thereby expanding the knowledge on the diversity of viruses infecting methanogens. One of these families was associated with the active hydrogenotrophic methanogenic archaea, while the other one seemed to target methanogens that were not assimilating the formate in the studied microcosms, which could only be evidenced thanks to the SIP analysis.

EXPERIMENTAL PROCEDURES

AD microcosm experiments and monitoring procedure

AD batch microcosms consisted in glass plasma bottles with a total volume of 500 ml. Either ^{13}C-labelled-formate or unlabelled formate was employed as the sole carbon source. A mineral solution suitable for AD (standard NF EN ISO 11734) was also added, as well as an inoculum prepared from a lab-scale digester fed with biowaste. To obtain the inoculum, the sampled anaerobic sludge was incubated in anaerobic conditions at 35°C for 20 days, for the majority of fermentable compounds to decompose. It was then centrifuged (10,000*g*, 20 min, 15°C), aliquoted and stored at −80°C until further use. The microcosms were hermetically sealed with an aluminium screw cap and a rubber septum. The headspaces were flushed with nitrogen gas (Linde). For each treatment (^{13}C-labelled or unlabelled formate), microcosms were prepared in triplicates with a working volume of 300 ml, using the mineral solution, 3 g of inoculum, and either 0.2 M ^{13}C-sodium formate (Cortecnet) or 0.2 M unlabelled sodium formate (Sigma Aldrich) (equivalent to 4.08 g).

During incubation, the following physicochemical parameters were monitored according to the protocols described in Puig-Castellví et al. (2022): biogas production, biogas composition, pH, isotopic composition of the biogas, volatile fatty acids (VFAs), total inorganic carbon (TIC), total organic carbon (TOC) and chemical oxygen demand.

To characterize the dynamics of microbial communities (archaea, bacteria and in particular their viruses), a pair of microcosms (one fed with ^{13}C-formate, one with unlabelled formate) was sacrificed at each of three different time points (days 8, 13 and 17), for cellular and viral DNA extraction. For each sacrificed microcosm, the liquid phase was centrifuged (8000*g*, 20 min, 15°C). The pellet was stored at −80°C for subsequent cellular DNA extraction, and the supernatant (containing VLPs) was collected and stored at 4°C for virion preparation.

Preparation and observation of VLPs

The supernatant resulting from the previous step was centrifuged once more (6000*g*, 20 min, 4°C) to further remove cells. The newly obtained supernatant was subjected to a three-step filtration with 1.2 and 0.8 μm polyethersulfone filters (Sartorius), and finally with 0.2 μm acetate cellulose filters. The filtrations were realized in 500 ml filtration units (Thermo Scientific) under vacuum. The final filtrate was centrifuged (40,000*g*, 3 h, 4°C) to pellet VLPs. The final pellet was suspended in 2 ml SM buffer (0.1 M NaCl, 0.1 M $MgSO_4$, 0.05 M Tris–HCl, pH 7.5) and stored at 4°C until subsequent analysis.

VLPs were observed by transmission electron microscopy (TEM) at MIMA2 MET - GABI, INRAE, AgroParisTech (78352 Jouy-en-Josas, France). Virion-containing solutions were adsorbed onto a carbon film membrane on a 300-mesh copper grid, stained with 1% uranyl acetate, dissolved in distilled water, and dried at room temperature. Grids were examined with a Hitachi HT7700 electron microscope operated at 80 kV (Elexience – France), and images were acquired with a charge-coupled device camera (AMT).

DNA extraction and quantification

Total cellular DNA was extracted with Qiagen DNeasy PowerSoil Kit, according to the manufacturer's instructions.

For VLPs, 360 μl of virion solution were treated with 2 U of Turbo DNase (2 U μl^{-1}, Ambion), and 22 μl of RNaseA (1 mg/ml, Ambion) for 1 h at 37°C, to remove contaminant non-encapsidated DNA. Inactivation buffer from the Ambion kit was added to stop the DNase activity. DNA extraction was then based on a classical phenol-chloroform method (Pickard, 2009), using 1.5 ml Phase lock gel light (VWR). The detailed procedure for viral DNA extraction is provided in Supplementary information (Section 1.1).

Concentrations of viral DNA were determined with Qubit dsDNA HS Assay Kit (Invitrogen), according to the manufacturer's instructions.

DNA ultracentrifugation

Ultracentrifugation on a CsCl gradient was applied to the total cellular DNAs to separate them according to their density, as described in Chapleur et al. (2016). Around 1 μg of DNA was employed for each sample. For each collected fraction, the DNA quantification was performed with Qubit dsDNA HS Assay Kit (Invitrogen). Based on DNA density profiles [Figure 1(D)], the fractions were pooled into three main groups (referred to as 'fraction pools' thereafter) and DNA was purified by glycogen-polyethylene glycol (PEG) 6000 precipitation, according to Neufeld et al. (2007).

The DNA of each fraction pool was amplified with RTG GenomiPhi™ V3 DNA Amplification Kit (illustra),

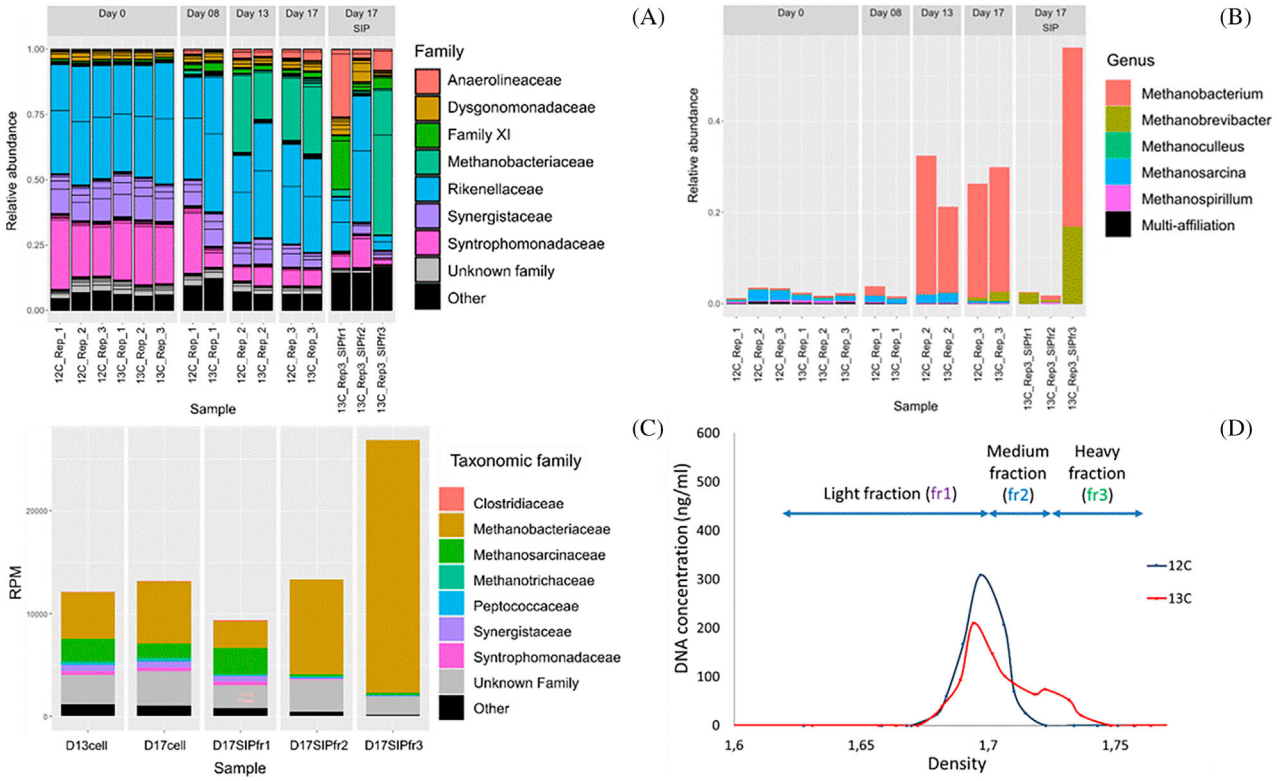


FIGURE 1 Microbial dynamics during AD incubation. (A) Microbial community composition of archaea and bacteria at the family level, based on 16S rRNA gene metabarcoding. (B) Composition of archaea only, at the genus level, based on 16S rRNA gene metabarcoding. (C) Taxonomic assignation (order level) of the microorganisms involved in methane metabolism, based on the functional analysis of the cellular shotgun metagenomes. (D) Density profiles of the cellular DNAs obtained from microcosms fed with unlabelled formate (blue) or with ^{13}C-formate (red) at day 17. The arrows labelled as fr1, fr2 and fr3 indicate how the DNA fractions were pooled in the case of labelled formate. fr1 contained only unlabelled DNA, whereas fr3 contained only ^{13}C-labelled DNA. fr2 likely contained a mix of ^{13}C-labelled and unlabelled DNA

according to the manufacturer's instructions. It was then purified by glycogen-PEG 6000 precipitation, and the concentration of the collected DNA was determined with Qubit dsDNA HS Assay Kit (Invitrogen).

16S rRNA gene metabarcoding

Archaeal and bacterial hypervariable regions V4–V5 of the 16S rRNA gene were amplified and then sequenced according to the protocol described in Poirier et al. (2016) and Madigou et al. (2019), with some modifications. The targeted region was amplified by PCR with fusion primers 515F (5′-Ion A adapter–Barcode–GTGYCAGCMGCCGCGGTA-3′) (Wang et al., 2007) and 928R (5′-Ion trP1 adapter–CCCCGYCAATTCMTTTRAGT-3′) (Wang & Qian, 2009). The detailed protocol for library preparation is provided in the Supplementary Information (Section 1.2). Sequencing was performed on an Ion Torrent Personal Genome Machine using Ion 316 Chip V2 (Life Technologies) and Ion PGM Hi-Q View Sequencing Kit (Life Technologies) according to the manufacturer's instructions. Sequencing data were processed with the Torrent Suite Software. All diversity analyses were performed with the R phyloseq package.

Shotgun metagenomic sequencing

Illumina sequencing was performed at the I2BC sequencing platform (University of Paris-Saclay, Gif-sur-Yvette, France). 250–500 ng of genomic DNA was fragmented (400 bp mean size) on a Covaris S220 sonicator. DNA fragments were end-repaired and dA-tailed (NEB#E7595), Illumina TruSeq adapters were ligated (NEB#E6040), and the PCR-free library fragments were purified using AMPure XP beads (Beckman Coulter). Final library quality was assessed on an Agilent Bioanalyzer 2100, using an Agilent High Sensitivity DNA Kit. Libraries were pooled in equimolar proportions and sequenced using paired-end 2 × 150 pb runs, on an Illumina NextSeq500 instrument, using NextSeq 500 High Output 300 cycles kit.

Demultiplexing was performed with bcl2fastq2 v2.18.12. Adapters were trimmed with Cutadapt v1.15, and only reads longer than 10 b were kept for further analysis.

Metagenomic pipe-line

The most generic steps of our pipeline were scripted as a snakemake (Köster & Rahmann, 2012) workflow (https://forgemia.inra.fr/cedric.midoux/workflow_metagenomics/-/tree/v21.04), and applied to both cellular and viral metagenomes. After a pre-processing step (adapter removal, and trimming according to quality scores and length with fastp), reads were assembled with metaSPADES (Nurk et al., 2016) (individual assembly by sample for cellular metagenomes and coassembly for metaviromes). Coding regions were predicted with Prodigal (Hyatt et al., 2010). Taxonomic affiliations of contigs and their predicted genes were respectively obtained with CAT (von Meijenfeldt et al., 2019) and kaiju (Menzel et al., 2016), against the NCBI nr database. Genes were annotated using comparison to NCBI nr database with Diamond (Buchfink et al., 2015). For each dataset, the cleaned reads were mapped to the assembled contigs using Bowtie2 (Langmead & Salzberg, 2012) and counted with samtools (Li et al., 2009). Details of versions and parameters for the snakemake workflow are available in the GitLab repository.

Several steps specifically dedicated to viral contig analysis and to host prediction were performed using homemade bash and python scripts. Metagenome-assembled genomes (MAGs) were constructed from assembled cellular metagenomic contigs with Metabat2 (Kang et al., 2019). MAG quality was improved with RefineM (Parks et al., 2017) and controlled with CheckM (Parks et al., 2015). Only MAGs with more than 60% completeness and less than 5% contamination were selected. Functional annotations of the cellular predicted genes were obtained with ghostKoala against KEGG database (Kanehisa et al., 2016). Concerning the viral metagenomes, their quality was assessed with ViromeQC (Zolfo et al., 2019) based on the trimmed reads. Viral genome detection was performed with VIBRANT (Kieft et al., 2020) and VirSorter2 (Guo et al., 2021), and the quality of viral contigs was assessed with CheckV (Nayfach et al., 2021). The predicted genes were further analysed by using HH-suite (Steinegger et al., 2019) against PHROGs (Terzian et al., 2021), a database dedicated to prokaryotic viruses.

To predict potential host for each viral contigs, two different methods were used. First, CRISPR spacers were detected in cellular metagenome contigs with CRISPRdetect (Biswas et al., 2016) and CRISPRCasFinder (Couvin et al., 2018). A non-redundant spacer database was built from the obtained spacer sequences. The viral contigs were subsequently aligned with BLASTn and SpacePHARER (Zhang et al., 2021) against this homemade database and a public spacer database (CRISPRCasdb spacer: https://crisprcas.i2bc.paris-saclay.fr/Home/DownloadFile?filename=spacer_34.zip), enabling host prediction based on an alignment-dependent method. Hosts were also predicted using the WIsH software (Galiez et al., 2017) to compare the genomic signatures of viral contigs to those of cellular MAGs.

After selecting viral contigs of interest as described in the result section, gene annotation was refined on HHpred server (Zimmermann et al., 2018) with default parameters, against four structural/domain databases: Pfam-A_v35, PDB_mmCIF70_12_Ot_2021, NCBI_Conserved_Domains(CD)_v3.18 and UniProt-SwissProt-viral70_3_Nov_2021.

The detailed parameters and versions for the python and bash scripts are presented in Supplementary Tables S2 and S3.

Bipartite network of viral contigs and protein orthologous groups

A bipartite network was built, where the two node classes were viral contigs or genomes on the one hand, and protein orthologous groups (OGs) on the other. Regarding 'viral nodes', most archaeal (pro)virus genomes available in public databases (RefSeqVirus and nr, 7th of October, 2021) or described in previous studies (Krupovič et al., 2010a; Philosof et al., 2017) were included, in addition to the viral contigs of interest. The final dataset consisted of 172 viral sequences, in total. All the proteins encoded in these sequences were categorized into OGs by a two-step procedure as described in Olo Ndela et al. (2021) and in the Supplementary Information (Section 1.2.2).

These OGs were functionally annotated by comparing their HMM profiles to those of the PHROGs database (Terzian et al., 2021) (annotation release v2), which contains well-annotated protein clusters of prokaryotic viruses. In addition, protein sequences were compared to RefSeqVirus using MMseqs (bit score ≥50), retaining only the best hit for each protein of an OG. These OGs were used as 'protein nodes' to construct the bipartite networks. The shared OGs and their functional prediction are listed in the Supplementary Table S4. All networks were computed using the igraph package of R (Csardi & Nepusz, 2005).

The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB46489 (https://www.ebi.ac.uk/ena/browser/view/PRJEB46489) (16S metabarcoding raw reads, shotgun metagenomics raw reads, annotated sequences of contigs C158, C889, C1359 and C1697).

RESULTS AND DISCUSSION

A significant proportion of formate was converted to methane through hydrogenotrophic methanogenesis

A total of six batch AD microcosms were established and monitored over time for a maximum of 17 days. They contained, as sole carbon source, either

unlabelled formate (three replicates) or ^{13}C-formate (three replicates). At each of days 8, 13 and 17, one pair of microcosms was sacrificed for metavirome analysis, one with unlabelled formate and one with ^{13}C-formate.

Chemical analyses (VFA, TIC, TOC) showed that both unlabelled and ^{13}C-labelled formate were consumed, and more than 85% of the initial quantity was metabolized after 17 days of incubation (Supplementary Figure S1). The bioconversion resulted in a pH increase from ~7.50 to ~9.20 on average, consistent with the methanogenesis equation mentioned above. Biogas was produced after a lag phase of about 5 days (Supplementary Figure S1) to reach final cumulated productions of ~100 normo-ml. Furthermore, reproducible dynamics were observed in microcosms, irrespective of the formate type. Methane (CH_4) was by far the dominant biogas component, followed by carbon dioxide (CO_2) as well as traces of hydrogen (H_2) and hydrogen sulfide (H_2S). Collectively, these results support the successful establishment of the AD process in the microcosms. More precisely, the isotopic composition of the biogas demonstrated the dominance of ^{13}C in the methane (>95%) produced by the methanogens in the microcosms fed with ^{13}C-formate and indicated the dominance of the hydrogenotrophic methanogenesis pathway (Whiticar et al., 1986) in the microcosms fed with ^{12}C-formate. For unlabelled substrates, this method relies on the abundance of the stable isotope ^{13}C in nature (1.1%) and on the difference in reaction rate between ^{12}C- and ^{13}C-containing substrate molecules. The results and detailed calculation of isotopic signatures are provided in the Supplementary Information (Section 2.1).

The proportion of methanogenic archaea increased over time and they were dominated by *Methanobacterium* species

To identify the methanogenic archaea and more broadly the microbial community composition, we applied 16S rRNA gene metabarcoding [Figure 1(A)]. The relative abundance of the dominant archaea (Methanobacteriaceae, in green) increased from 1%–3% at day 0 to 26%–30% at day 17. Consistently, a functional analysis based on the cellular shotgun metagenomic data (Supplementary Figure S2) showed the dominance of methane metabolism, in relative abundance, at day 17. The KEGG category 'methane metabolism' includes methanogenesis, methane oxidation and metabolisms related to intermediate molecules of both these pathways. Among the microbial groups involved in methane metabolism [Figure 1(C)], archaea were the main actors (Methanobacteriaceae in brown, Methanosarcinacea in green and Methanotrichaceae in cyan). Similar proportions of archaea were observed in the metagenomic dataset (Supplementary Figure S3). These results confirmed that the microbial communities were enriched in methanogenic archaea during the incubation.

The detected archaea were mostly methanogens [-Figure 1(B)]. At day 0, *Methanosarcina* (Methanosarcinales order) was dominant. However, at the end of the incubation, at day 17, the genus *Methanobacterium* became dominant, followed by *Methanobrevibacter* (both from the order Methanobacteriales). It can be assumed that Methanobacteriales members were selected during the incubation since they are able to grow on formate through the hydrogenotrophic pathway, and since they likely outcompeted Methanosarcinales methanogens, due to their faster growth rate: the doubling time is generally lower than 1 h in Methanobacteriales, compared to more than 10 h in Methanosarcinales (Thauer et al., 2008).

Concerning bacterial families present in these systems, a notable decrease in relative abundance was observed over time for *Synergistaceae* (phylum Synergistetes, from ~15% down to ~7%) and *Syntrophomonadaceae* (phylum Firmicutes, from ~24% down to ~7%) [Figure 1(A)]. *Synergistaceae* members generally consume amino acids to generate short-chain fatty acid (He et al., 2018), whereas *Syntrophomonadaceae* members are acetogens (Si et al., 2016). For both families, the decrease is understandable due to the lack of adequate substrate (proteins and short-chain fatty acids). A notable increase in relative abundance was observed at day 17 for members of *Anaerolineaceae* (phylum Chloroflexi, from >1% up to ~4%), which are generally described as fermentatives or acetogens (Liang et al., 2015; Si et al., 2016), and some of them were previously shown to form syntrophic associations with methanogens (Lei et al., 2018). In the present experiment, *Anaerolineaceae* bacteria may degrade formate and/or play a role in electron transfer (Wang et al., 2021), in partnership with hydrogenotrophic methanogens.

For the six microcosms, the profiles of cellular DNA concentrations in function of their mass density were established (Supplementary Figure S4). For microcosms fed with ^{13}C-formate at day 17 [Figure 1(D), red line], a second peak, of denser DNA, was visible in the profile, indicating that the labelled substrate was assimilated by some of the microorganisms. The separation of DNA in CsCl gradient depends not only on the mass of the isotope but also on the GC content of the DNA (Eason & Campbell, 1978). Hence, we pooled the DNA into three distinct fractions, according to their density. The first fraction (fr1, density <1.705) corresponded exclusively to non-labelled DNA. The second fraction (fr2, 1.705 < density <1.738) possibly contained a mix of unlabelled and ^{13}C-labelled DNA. Finally, the third fraction (fr3, density >1.738) contained exclusively ^{13}C-labelled DNA. The 16S rRNA gene metabarcoding applied to these fractions showed that archaea reached

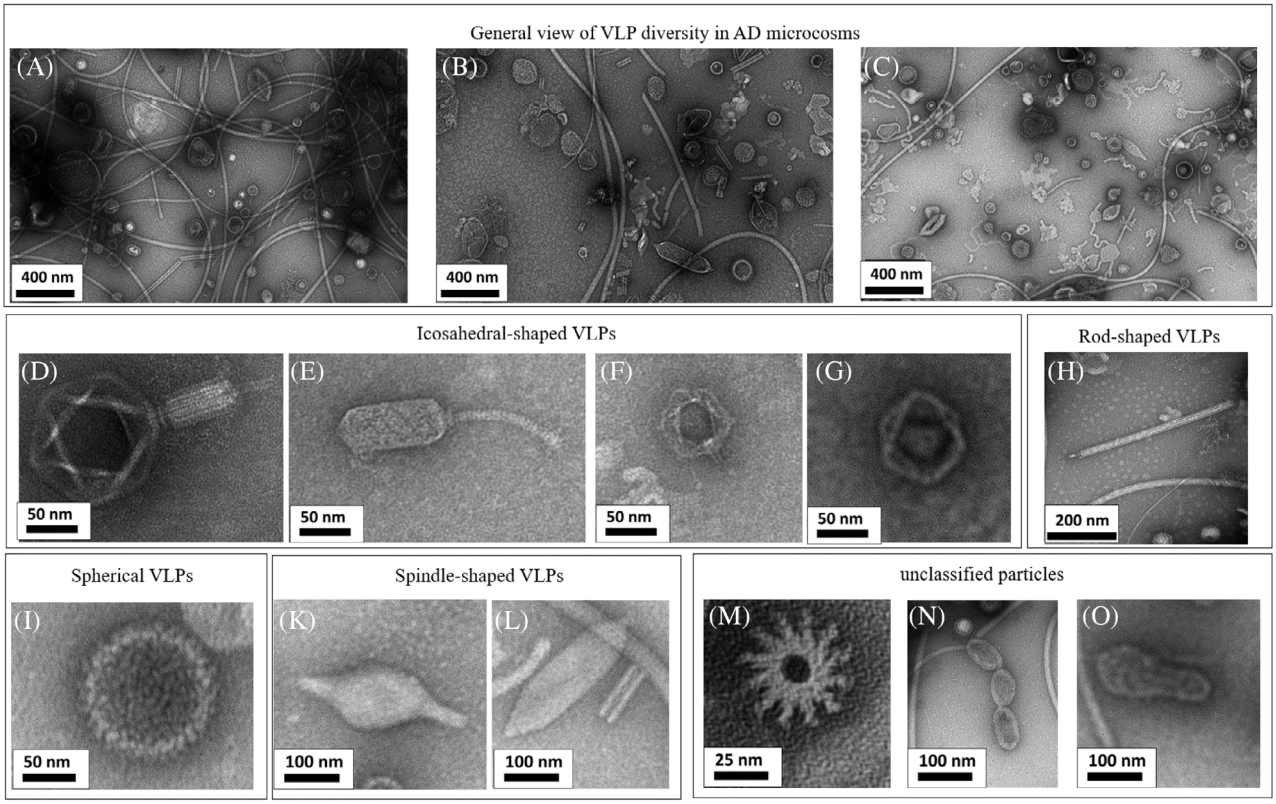


FIGURE 2 Morphotypic diversity of VLPs in different microcosms, observed by TEM. One representative sample is shown for each incubation time point (A: day 8, B: day 13, C: day 17)

their highest proportions (56%) in fraction fr3 at day 17 [-Figure 1(B)], whereas their proportions in fractions fr1 and fr2 were lower than 3%. Consistently, the highest abundance of genes involved in methane metabolism was in fraction fr3 at day 17, based on metabolic pathway analysis from the shotgun metagenomics data [-Figure 1(C)]. These observations confirmed that the ^{13}C-formate incorporated into the microbial biomass was consumed mostly by methanogenic archaea of the genera *Methanobacterium* and *Methanobrevibacter*.

TEM evidenced the presence of icosahedral and spindle-shaped VLPs

TEM observations revealed a great diversity of VLPs in the microcosms (Figure 2), especially after 13 and 17 days of incubation [Figure 2(A,C)]. Besides cosmopolitan morphotypes common to the domains Bacteria and Archaea, uncommon and especially archaea-specific morphotypes were also observed.

Cosmopolitan morphotypes included VLPs with a head-tailed morphology, typical of the class *Caudoviricetes*, as well as icosahedral tailless particles [Figure 2(G)]. The latter could originate from tailless icosahedral viruses or from head-tailed viruses with a broken tail. In both samples, these VLPs were the most abundant and presented a large diversity: myovirus-like [i.e. icosahedral capsids with contractile tails; Figure 2(D)], siphovirus-like [i.e. icosahedral capsids with long non-contractile tails; Figure 2(E)] and also podovirus-like [i.e. icosahedral capsids with short tails; Figure 2(F)] morphotypes were observed, with capsid diameters ranging from 50 to 200 nm.

Less common viral morphotypes were detected in the microcosms at days 13 and 17, such as rod-shaped [-Figure 2(H)], spherical [Figure 2(I)] and spindle-shaped [-Figure 2(K–L)]. Spindle-shaped morphotypes, which are specific to archaeal viruses, have been commonly observed in extreme geothermal and hypersaline environments (Krupovic et al., 2014), but have also been reported in moderate ones, such as freshwater and marine habitats (Borrel et al., 2012; Kim et al., 2019). Interestingly, the presence of spindle-shaped VLPs has also been reported in AD plants (Calusinska et al., 2016), suggesting that the hosts of these viruses could possibly be involved in the AD process.

Particles with unique morphotypes were also identified [Figure 2(M–O)]. Some of them had a chain structure with multiple of two or three monomers and their size ranged from 100 to 400 nm. Moreover, other particles appeared as round-shaped and less than 50 nm, with a hollow star-like structure [Figure 2(M)]. The nature of these particles is unclear.

TABLE 1 Overview of the shotgun metagenome datasets

	Cellular metagenomes (individual assemblies)					Viral metagenomes (co-assembly)	
	D13cell	**D17cell**	**D17SIPfr1**	**D17SIPfr2**	**D17SIPfr3**	**D13vir**	**D17vir**
Number of raw reads (millions)	85.68	63.76	75.41	78.78	65.54	38.92	46.48
Reads obtained after trimming (%)	98.39	98.08	97.80	98.35	98.40	98.41	98.61
Number of contigs (≥1 kb)	77,032	68,382	58,308	33,123	21,159	23,016	
Number of contigs (≥3 kb)	18,093	15,244	14,728	9432	5108	5571	
Max contig length (b)	622,910	622,910	449,342	388,850	389,221	372,273	
Contigs with taxonomic affiliation (%)	96.62	96.67	96.93	97.32	97.22	87.59	
Reads mapped to contigs (%)	89.35	87.74	89.03	93.17	93.73	87.20	79.85
Number of MAGs	81	67	75	55	28	–	
Number of archaeal MAGs	8	10	8	7	5	–	
Number of selected MAGs (≥60% completeness and ≤5% contamination)	81 (including 17 from archaea)						

Overview of the shotgun metagenomics datasets and of the host prediction approaches

Shotgun metagenomic sequencing was applied to seven selected DNA extracts. Four of the extracts were from the total cellular DNA and total viral DNA from samples collected at days 13 and 17 (D13cell, D13vir, D17cell, D17vir), selected due to their enrichment in methanogens observed in the metabarcoding analysis (Figure 1) and the presence of interesting VLP morphotypes (Figure 2). The three cellular DNA fractions obtained after density gradient centrifugation at day 17 were also sequenced (D17SIPfr1, D17SIPfr2, D17SIPfr3) [Figure 1(D)]. The purpose was primarily to analyse the metaviromes, and to use the cellular metagenomes for building specific databases for host prediction.

Based on classical metrics (Table 1), the sequencing data and the assemblies were of satisfactory quality. In the case of metaviromes, ViromeQC analysis indicated moderate contamination rates of 7.6% and 8.8% for D13vir and D17vir, respectively. Among 23,016 assembled contigs longer than 1 kb, 1927 were exclusively detected in D17vir, suggesting that most of the corresponding viruses were selected during the late stages of the incubation. Among the 87.59% of contigs that obtained a taxonomic annotation, 98.43% were affiliated to prokaryotes, which can be explained primarily by database biases: prokaryotes were more represented than their viruses in the NCBI nr database used for the taxonomic assignment, and many viral genes could therefore have their best match in prokaryotic genomes, in particular in proviruses. Such bias is a strong limitation for the identification of viruses and their taxonomic classification. Yet, it can bring information on the possible hosts of the viral contigs.

Viral genome detection was performed with CheckV, VIBRANT and VirSorter2. All these tools rely on protein similarity profiles and both latter detect protein profiles by machine learning. CheckV, VirSorter2 and VIBRANT identified 3898, 3968 and 3904 viral genomes (6484 distinct genomes in total), respectively, with most of the genomes being linear, double stranded and predicted to belong to virulent viruses. These numbers were low compared to the number of contigs longer than 1 kb, likely reflecting the fact that shortest contigs contain insufficient information for confident virus identification.

For cellular metagenomes, a taxonomic affiliation was obtained for more than 96% of the contigs, suggesting that they would constitute a good-quality database of host sequences. MAGs were reconstructed by binning of the cellular contigs: in total, 81 MAGs (completeness >60% and contamination <5%) were selected as reference hosts, including 17 assigned to the domain Archaea.

For host prediction, two complementary methods were used. The first one relied on matching of CRISPR spacers to the viral contigs. Spacers are short sequences (26–72 bp) (Makarova et al., 2011) of plasmid or viral origin, which are integrated into the host genomes by CRISPR-Cas (Nasko et al., 2019), adaptive immunity systems identified in 85% of Archaea and 40% of Bacteria (Makarova et al., 2020). The second approach relied on signatures, as most prokaryotic viruses have a genome k-mer composition similar to the one of their hosts, due to their co-evolution (Edwards et al., 2016). Similar genomic signatures were searched between viral contigs and the 81 microbial MAGs as potential hosts, using WIsH (Galiez et al., 2017). For contigs longer than 3 kb, the spacer-based method revealed a total of 375 contigs with a predicted host, whereas the signature-based method

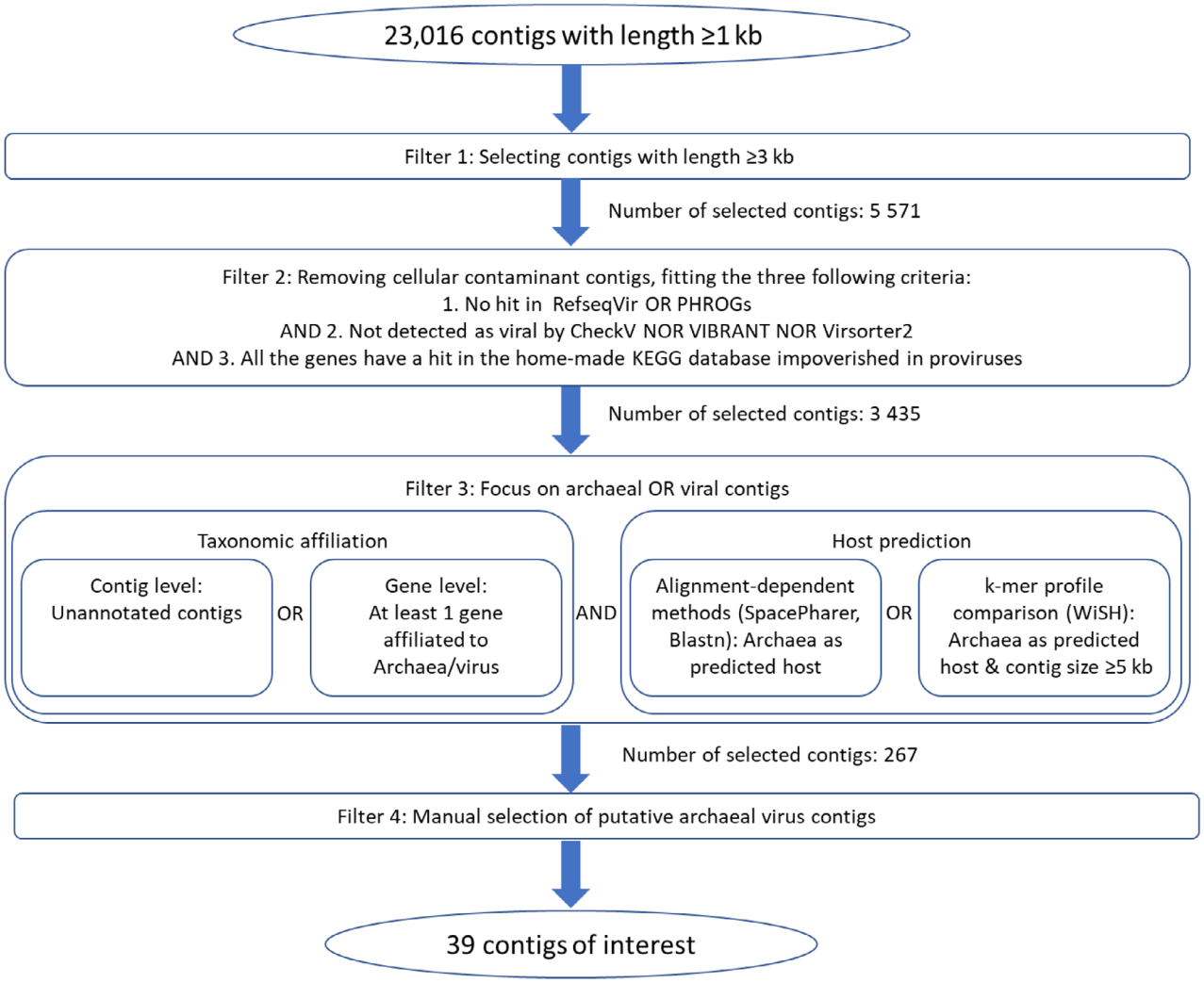


FIGURE 3 Strategy for the selection of contigs of interest, likely originating from viruses of methanogens

against our homemade MAG database predicted an acceptable host (p-value ≤0.05) for 3239 contigs. Coupling of these methods resulted in a total of 3466 contigs with a predicted host, showing their complementarity to improve the capability and accuracy of *in silico* host prediction.

The *Caudoviricetes* class dominated among the 39 contigs likely originating from viruses of methanogens

As our aim was to identify contigs of viruses infecting methanogenic archaea, we applied successive stringent filters and manual curation, relying on the integration of results from complementary bioinformatic analyses (Figure 3). A first filter based on contig length was applied to eliminate contigs which would contain only limited information. The second filter aimed at removing an important fraction of the cellular contaminants. In the third step, contigs possibly originating from archaea or archaeal viruses were selected. Finally, the contigs obtained at this stage were analysed manually, to remove most remaining contaminant contigs (typically originating either from cells or from bacterial viruses).

A total of 39 contigs were thereby selected as contigs of interest for further analysis, as putatively originating from archaeal viruses. Due to the limited accuracy of some host prediction tools [e.g. of the order of 70%–80% at the phylum level (Galiez et al., 2017)], this selection of contigs still possibly contained a few contaminants. The selected contigs ranged in lengths from 4.2 to 53 kb (median: 9.4 kb) and the reads per kilobase per million mapped reads (RPKM) values from 0.07 to 468.27 for D13vir, and from 0.20 to 229.40 for D17vir. Nine of these contigs could not be assigned to any known taxon (either cellular or viral), suggesting them to be previously uncharacterized. Two contigs, C1661 and C1697, had best-predicted hosts as MAGs

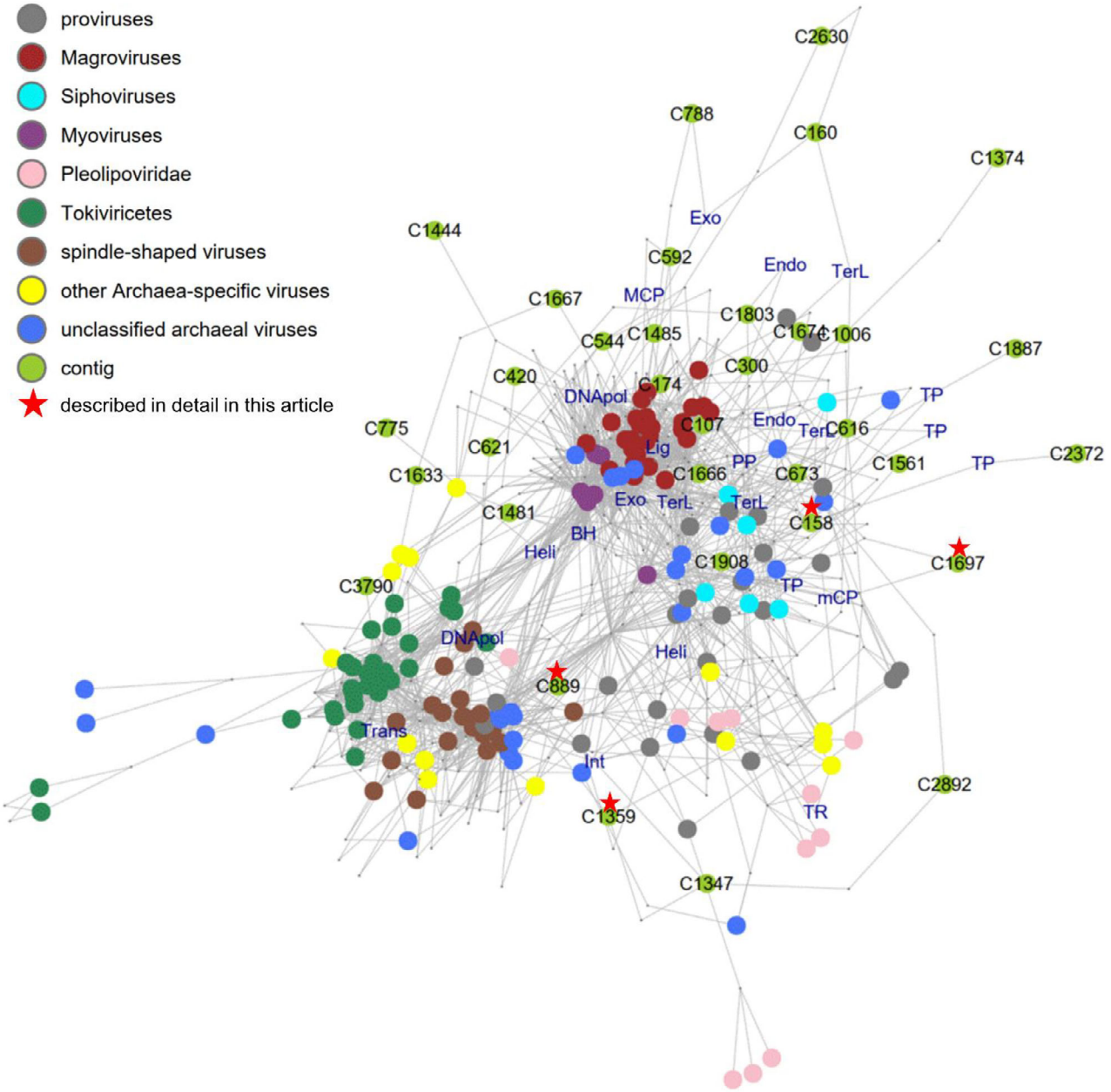


FIGURE 4 Bipartite network of known archaeal viruses and of the 39 contigs of interest, and of protein orthologous groups (OGs). Most of the contigs of interest originate from archaeal viruses. The genomes and contigs are represented as circles coloured according to the legend shown in the figure, and OGs are denoted by the intersections of edges. Some shared OGs with a functional annotation are represented. TerL: Terminase large subunit, MCP: major capsid protein, mCP: minor capsid protein, Exo: exonuclease, TP: tail protein, PP: portal protein, Lig: ligase, DNApol, DNA polymerase, BH: baseplate hub protein, Heli: helicase, Trans: transposase, TR transcription regulator

from the unlabelled metagenome D17SIPfr1, suggesting that they represent viruses infecting inactive and minor methanogens in our microcosms. Besides, 34 were predicted as virulent by VIBRANT and only one, C1485, was predicted as temperate (Supplementary Figure S5, Supplementary Table S5). Genome scaffold quality results were obtained through three different bioinformatic tools: VIBRANT, CheckV and VirSorter2: 35 contigs had low or medium quality. However, given that all of these tools were developed based on the viral databases overwhelmingly dominated by bacterial *Caudoviricetes*, their accuracy on datasets including novel archaeal viruses, especially those with small or medium-sized genomes, is not to be expected. All the results from the bioinformatics tools are available for this set of 39 contigs, in Supplementary Table S5.

A bipartite network was built (Figure 4) to evaluate the similarity between the contigs of interest and the archaeal (pro)viruses described in the literature. Such

networks enable to represent complex systems comprising two distinct classes of components (nodes) (Iranzo et al., 2016b). Here, the two classes of node correspond to viral contigs or genomes on the one hand, and to protein OGs on the other. Two viral nodes can be connected only indirectly, through shared OG nodes. Previously published archaeal pro(viral) genomes were labelled according to their taxonomic affiliation, when established. The top half of the network contained viral genomes from (pro)viruses with a cosmopolitan head-tailed morphotype (siphovirus, myovirus, magrovirus, provirus). In the bottom half, the presence of various viral families specific to archaea was observed. Such a network topology is consistent with the spatial distribution in archaeal virus networks previously described in the literature (Iranzo et al., 2016a). Moreover, head–tail viruses (class *Caudoviricetes*) are known to be highly mosaic (Krupovič et al., 2010), hence generating tightly interconnected networks.

Out of 39 contigs of interest (green), 24 shared at least one OG with (pro)viruses belonging to the class *Caudoviricetes*, including virion morphogenesis proteins, such as terminase large subunit (TerL), various tail proteins, major/minor capsid proteins (MCP/mCP) and baseplate hub proteins (BH).

To further explore the organization of the network and relationships among the 39 contigs of interest and archaeal (pro)viruses, they were clustered according to the presence/absence of shared genes (see Experimental procedures). Consistent with the network analysis, the two main viral clusters (VC) were related to the class *Caudoviricetes* (Supplementary Figure S6). The first cluster, VC4, included four contigs, C1803, C174, C300 and C616, and was held together through OGs related to the capsid formation and packaging module, and/or DNA, RNA and nucleotide metabolism (Supplementary Figure S6). Moreover, in the longest contig of VC4, C174, a sheath protein was detected, also suggesting its affiliation to viruses with contractile tails (myovirus-like morphology) (Fokine & Rossmann, 2014). For all the contigs in this cluster, only WIsH predicted an archaeal host, whereas the other tools employed showed either a bacterial host or no identified one. To ascertain the host assignment, the taxonomic and functional annotations of each gene from the cluster were examined: no single gene was assigned to Archaea or annotated as an archaeal protein, while many were related to Bacteria, suggesting that bacterial hosts were more probable. It highlights the importance of relying on different complementary tools for more accurate host prediction.

The second cluster, VC10, included five contigs, C1666, C1908, C158, C673 and C1006. Several OGs shared among members of the cluster were related to various viral functional modules: virion morphogenesis (including capsid and tail formation, and genome packaging), DNA/RNA and nucleotide metabolism, as well as integration and excision. In particular, C158 seems to be closely related to C673 and C1908 as they share respectively six and two OGs. VC10 was located near the siphovirus nodes (cyan) in the network. Importantly, the taxonomic affiliation of these contigs and their host prediction by different tools were consistent with each other and confirmed that they originated from archaeal viruses, unlike contigs from VC4.

Contigs located on the periphery of the network shared very few OGs with other known viruses (only one or two OGs per contig). Moreover, those shared OGs were often uncharacterized or hypothetical proteins. Several contigs shared annotated OGs belonging to viral families specific to archaea, and they could have originated from previously uncharacterized viruses of methanogens. For example, C775 shared an OG annotated as helicase with eight genomes belonging to the family *Lipothrixviridae*. However, only the presence of signature genes, that is characteristic of particular virus groups, such as those encoding major structural proteins, can provide a reliable taxonomic affiliation of contigs, as it is the case for VC4 and VC10 contigs. For contigs outside of these two VC, such signature genes were identified only for two contigs (C1359 and C1697). Indeed, a structural protein typical of archaeal spindle-shaped viruses (Krupovic et al., 2014) was identified in these two contigs, suggesting that they represent new families of spindle-shaped viruses, as described in more detail in the next section.

Viruses infecting active methanogens during the incubation were identified

Among the 39 viral contigs of interest, several had predicted hosts corresponding to the dominant and active methanogens in the studied microcosms, namely, those from the order Methanobacteriales (*Methanobacterium* and *Methanobrevibacter* genera).

In particular, contig C158 (belonging to VC10) was predicted to infect *Methanobacterium* species, according to its taxonomic affiliation and the signature-based host prediction. *Methanobacterium* was the most abundant archaeal genus in the studied microcosms. C158 has a length of 42,490 bp and contains 49 predicted genes. Consistent with the abundance of the predicted host, C158 was the most abundant of the 39 contigs in metaviromes at days 13 and 17, with RPKM values of 468 and 229, respectively. Furthermore, C158 is a complete dsDNA circular genome, from a head-tailed virus. All functional modules typical for the class *Caudoviricetes* were identified in this contig (Figure 5), such as those required for virion morphogenesis (HK97-like MCP, mCP, tail proteins, terminase subunits, capsid maturation protease) and several

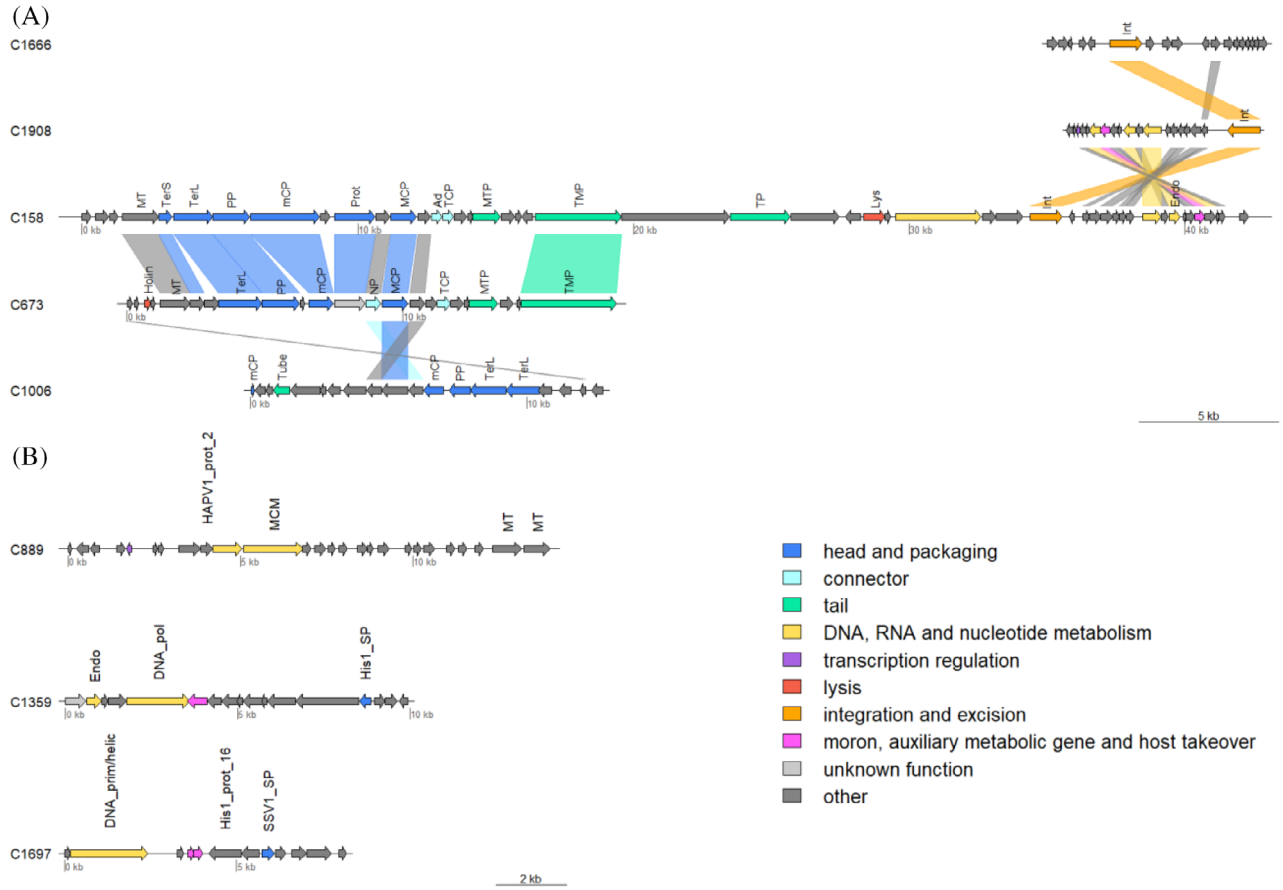


FIGURE 5 Genome map for selected contigs of interest. (A) Contigs from VC10 affiliated to class *Caudoviricetes*, including C158 (proposed new family *Sperovirida*e). (B) Viral contigs not related to head-tailed viruses, including predicted spindle-shaped viruses. Genes were annotated with MMseq against PHROGs and with hhpred against four other databases (see Experimental procedures). Genes are represented as arrows. The main functional modules are indicated by colours. Abbreviations of core viral proteins: Ad, adaptor protein; BP, baseplate protein; BH, baseplate hub protein; BS, baseplate spike protein; BW, baseplate wedge protein; DNA_pol, DNA polymerase; DNA_prim, DNA primase; DNA_prim/helic, DNA primase/helicase; Endo, endonuclease; HAPV1_prot2, similar to protein 2 from *Halorubrum* pleomorphic virus 1; His1_prot_16, similar to protein 16 from *Haloarcula hispanica* virus 1; His1_SP, His1-like major capsid protein; Holin, holin; Int, integrase; Lys, endolysin; MCM, MCM helicase; mCP, minor coat protein; MCP, major capsid protein; MT, methyl transferase; MTP, major tail protein; NP, preneck appendage protein; NT, tRNA nucleotidyltransferase; PP, portal protein; Prot, capsid maturation protease; Sc, scaffolding; Sh, tail sheath; SSP1_SP, similar to the main structural protein of *Shigella* phage SSP1; SSV1_SP, SSV1-like major capsid protein; TCP, tail completion protein; TerL, terminase large subunit; TerS, terminase small subunit; TMP, tail tape measure protein; TP, tail protein; Tube, tail tube

enzymes necessary for the life cycle, such as an intracellular proteinase inhibitor and an integrase, this latter suggesting a temperate lifestyle. According to its proximity to siphoviruses in the network (Figure 4) and VIRFAM analysis (Lopes et al., 2014), C158 is likely to have a siphovirus-like morphology, that is long non-contractile tail. Considering the lack of very strong similarity with previously characterized viruses, it likely represents a new family of archaeal viruses within the *Caudoviricetes*. We suggest the name *Sperovirida*e for this new viral family.

The second-most abundant contig of interest was C1359, one of the two contigs predicted to have a spindle-shaped morphology, with RPKM values at days 13 and 17 of 285 and 183, respectively. Surprisingly, this 9936 bp-long contig (17 detected genes) was predicted to infect *Methanoculleus* sp. (order Methanomicrobiales), which was present at very low abundances over the incubation time (<0.5% of the archaea based on shotgun metagenomic data). This host prediction was based on spacer alignment (SpacePHARER) against a public spacer database. By contrast, WIsH and spacer alignment with blastn did not yield any high-confidence predicted host. Based on the archaeal community composition (see section 3.1.2), this is likely a false prediction, despite a highly significant *p*-value (2.03×10^{-7}). Nevertheless, C1359 gene content confirmed the probable archaeal nature of its host, and its high abundance suggests that it infected a dominant methanogen (order Methanobacteriales). Indeed, it encoded a protein showing a significant similarity with the major capsid proteins of two spindle-shaped archaeal viruses, the halophilic *Haloarcula hispanica* virus His1 (Bath & Dyall-Smith, 1998)

and the hyperthermophilic *Sulfolobus shibatae* virus SSV1 (Palm et al., 1991) (probability: 98.84% and 98.36% respectively). Moreover, spindle-shaped viruses are specific to archaea (Krupovic et al., 2018; Pina et al., 2011; Prangishvili et al., 2017; Snyder et al., 2015), excluding the possibility that the virus infects bacteria. This predicted morphotype is consistent with the presence of spindle-shaped VLPs observed by TEM [Figure 2(K–L)] and reported in AD ecosystems (Calusinska et al., 2016). Due to its limited similarity with previously characterized viruses, C1359 likely represents a new viral family, and corresponds to the first spindle-shaped virus reported in association with the order Methanobacteriales.

Viruses infecting minor methanogens in the formate incubation microcosms were identified

As mentioned above, methanogens of the orders Methanomicrobiales and Methanosarcinales were present in the studied microcosms at a low abundance (<5%). Nevertheless, previously uncharacterized viral genomes possibly infecting these methanogens were identified.

Contig C889 has a length of 14,120 bp and contains 26 predicted genes. Its RPKM values at days 13 and 17 were 2.3 and 3.3, respectively. Based on its taxonomic affiliation, this contig seemed to originate from a virus infecting a Methanomicrobiales host. In this contig, a protein similar to replicative minichromosome maintenance helicases was detected. This is not a structural protein but it has been identified in several families of archaeal viruses and archaeal plasmids (Krupovič et al., 2010a), strongly supporting an archaeal host. No structural protein could be identified in C889, suggesting either a partial genome or a new type of virus.

Contig C1697 has a length of 8207 bp, contains 11 predicted genes, and had RPKM values of 4.1 at day 13 and 8.4 at day 17. Based on its taxonomic affiliation and on host prediction with WIsH, this contig originates from a *Methanosarcina* virus. It encodes one protein showing significant profile similarity with the major capsid protein of SSV1 (*Fuselloviridae*) (probability: 94.92%), suggesting a spindle-shaped virion morphology. C1697 did not show pronounced similarity with C1359 (which also had a predicted spindle-shaped morphotype) or with known spindle-shaped viruses, suggesting that it could represent a second new family of spindle-shaped archaeal viruses. Notably, however, similar to spindle-shaped halspivirus His1 (Bath & Dyall-Smith, 1998) and thaspivirus *Nitrosopumilus spindle-shaped virus 1* NSV1 (Kim et al., 2019), C1697 encodes a protein-primed family B DNA polymerase, suggesting that the genome of this virus is linear.

CONCLUDING REMARKS

Predicting the hosts of viruses is a major bottleneck in microbial ecology. In recent years, besides the classic culture-dependent approaches, several promising methods have been developed to identify the hosts for viruses within complex microbial communities, such as PhageFISH (Allers et al., 2013; Barrero-Canosa & Moraru, 2019), Meta3C (Marbouty et al., 2014), viral tagging (Deng et al., 2014) or epicPCR (Sakowski et al., 2021; Spencer et al., 2016). These single cell level methods generated important new knowledge, but they also suffer from limitations, either due to the requirement of a pre-existing knowledge on the viral genome sequence, of the capability to cultivate the host, or due to their complexity. Several purely bioinformatic approaches of host prediction have been also developed (Coclet & Roux, 2021; Edwards et al., 2016). However, the accuracy of *in silico* methods is limited, in most cases, by the lack of microbial genomes closely related to that of the true host. Viral host prediction, therefore, remains challenging for metagenomics studies.

The original experimental approach developed in the present study is not strictly speaking a method for the identification of viral hosts but it still presents some advantages in this perspective. Relying on SIP enabled us to discriminate between the active microorganisms involved in formate metabolism and the other ones. Although this substrate was not consumed exclusively by methanogens, the use of formate strongly enriched the community in methanogens and their associated viruses, resulting in a simplified microbial community to study and, possibly, an improved MAG assembly. Coupling the results from different bioinformatic methods (taxonomic assignation of the contigs, alignment- and signature-based host prediction), we were thus able to detect several previously uncharacterized genomes of viruses infecting methanogens, with high accuracy, and to determine whether they targeted methanogens actively involved in the formate bioconversion. It was especially possible thanks to the establishment of specific host databases, through the shotgun sequencing of cellular metagenomes. Interestingly, one of them likely corresponds to a new archaeal virus family within the *Caudoviricetes* (proposed name *Sperovirida*e), and two others to putative new families of spindle-shaped archaeal viruses. These results significantly expand the knowledge on the diversity of viruses of methanogens, since no spindle-shaped virus has been reported until now for the orders Methanobacteriales and Methanosarcinales. It illustrates the complementarity between SIP, metagenomics and specific bioinformatic tools for host-virus analysis in complex microbial communities. Our approach can be viewed as complementary to the available experimental methods for host identification, and could likely be combined with most of them. In particular, PhageFISH or epicPCR would be nicely

complementary as they require at least a partial knowledge of the viral sequence, and they would enable to fully confirm the host identity.

To sum up, we proposed an original coupling between SIP, an experimental method, and metagenomic analyses with complementary bioinformatic tools to identify viruses of methanogenic archaea within anaerobic digestion microcosms. It enabled successful enrichment of the microbial community in methanogenic archaea, and their labelling with ^{13}C. The *in silico* approach we developed led to the identification of several dozens of contigs predicted to originate from viruses infecting methanogenic archaea. Their analyses through gene-sharing network and comparative genomics highlighted the dominance of the *Caudoviricetes* class, with the discovery of a previously uncharacterized siphovirus infecting Methanobacteriales hosts and belonging to a new suggested viral family, *Sperovirida*e. It also led to the discovery of two spindle-shaped viruses representing two new putative families, not previously reported for the orders Methanobacteriales and Methanosarcinales. Our results significantly expand the knowledge on the diversity of viruses of methanogens and reinforce the notion of the wide environmental and phylogenetic distribution of spindle-shaped viruses in Archaea (Krupovic et al., 2020). Our original experimental approach enables to identify viruses infecting key functional groups contributing to biogeochemical fluxes in communities of uncultured microbes. It can be invaluable for the study of viruses infecting metabolically active microorganisms in virtually any type of complex microbial community.

ACKNOWLEDGEMENTS

We are grateful to Chrystelle Bureau for her technical support on 16S rDNA metabarcoding, and to Angeline Guenne and Nadine Derlet for their technical assistance in analytical chemistry, in INRAE-PROSE. We also would like to acknowledge Christine Longin of the INRAE-MIMA2, platform for amazing TEM observations. This work has benefited from the facilities and expertise of MIMA2 MET – GABI, INRA, AgroParisTech, 78352 Jouy-en-Josas, France. We also acknowledge the high-throughput sequencing facility of I2BC (University of Paris-Saclay, Gif-sur-Yvette, France) for its sequencing. This work was supported by Agence Nationale de la Recherche, France (ANR-17-CE05-0011, project VIRAME). MK was supported by Agence Nationale de la Recherche grant ANR-20-CE20-0009.

CONFLICT OF INTEREST

The authors declared to have no conflict of interest.

ORCID

Vuong Quoc Hoang Ngo https://orcid.org/0000-0001-6746-5521
Ariane Bize https://orcid.org/0000-0003-4023-8665

REFERENCES

Ahring, B.K. (2003) Perspectives for anaerobic digestion. In: Ahring, B.K., Angelidaki, I., de Macario, E.C., Gavala, H.N., Hofman-Bang, J., Macario, A.J.L. et al. (Eds.) *Biomethanation I. Advances in biochemical engineering/biotechnology*. Berlin, Heidelberg: Springer, pp. 1–30.

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J. et al. (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology*, 15, 2306–2318.

Baquero, D.P., Liu, Y., Wang, F., Egelman, E.H., Prangishvili, D. & Krupovic, M. (2020) Chapter four - structure and assembly of archaeal viruses. In: Kielian, M., Mettenleiter, T.C. & Roossinck, M.J. (Eds.) *Advances in virus research. Virus assembly and exit pathways*. Cambridge: MA: Academic Press, pp. 127–164.

Barrero-Canosa, J. & Moraru, C. (2019) PhageFISH for monitoring phage infections at single cell level. *Methods in Molecular Biology*, 1898, 1–26.

Bath, C. & Dyall-Smith, M.L. (1998) His1, an archaeal virus of the Fuselloviridae family that infects Haloarcula hispanica. *Journal of Virology*, 72, 9392–9395.

Biswas, A., Staals, R.H.J., Morales, S.E., Fineran, P.C. & Brown, C. M. (2016) CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics*, 17, 356.

Borrel, G., Colombet, J., Robin, A., Lehours, A.-C., Prangishvili, D. & Sime-Ngando, T. (2012) Unexpected and novel putative viruses in the sediments of a deep-dark permanently anoxic freshwater habitat. *The ISME Journal*, 6, 2119–2127.

Buchfink, B., Xie, C. & Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59–60.

Calusinska, M., Marynowska, M., Goux, X., Lentzen, E. & Delfosse, P. (2016) Analysis of dsDNA and RNA viromes in methanogenic digesters reveals novel viral genetic diversity. *Environmental Microbiology*, 18, 1162–1175.

Chapleur, O., Mazeas, L., Godon, J.-J. & Bouchez, T. (2016) Asymmetrical response of anaerobic digestion microbiota to temperature changes. *Applied Microbiology and Biotechnology*, 100, 1445–1457.

Coclet, C. & Roux, S. (2021) Global overview and major challenges of host prediction methods for uncultivated phages. *Current Opinion in Virology*, 49, 117–126.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B. et al. (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, 46, W246–W251.

Csardi, G. & Nepusz, T. (2005) The Igraph software package for complex network research. *InterJournal Complex Systems*, 695, 1–9.

Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P. et al. (2014) Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. *Nature*, 513, 242–245.

Eason, R. & Campbell, A.M. (1978) 8 - Analytical ultracentrifugation. In: Birnie, G.D. & Rickwood, D. (Eds.) *Centrifugal separations in molecular and cell biology*. London, UK: Butterworth-Heinemann, pp. 251–287.

Edwards, R.A., McNair, K., Faust, K., Raes, J. & Dutilh, B.E. (2016) Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiology Reviews*, 40, 258–272.

Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P. et al. (2019) An evolving view of methane metabolism in the archaea. *Nature Reviews Microbiology*, 1, 219–232.

Fokine, A. & Rossmann, M.G. (2014) Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage*, 4, e28281.

Galiez, C., Siebert, M., Enault, F., Vincent, J. & Söding, J. (2017) WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics*, 33, 3113–3114.

Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O. et al. (2021) VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, 9, 37.

He, J., Wang, X., Yin, X., Li, Q., Li, X., Zhang, Y. et al. (2018) Insights into biomethane production and microbial community succession during semi-continuous anaerobic digestion of waste cooking oil under different organic loading rates. *AMB Express*, 8, 92.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W. & Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 119.

Iranzo, J., Koonin, E.V., Prangishvili, D. & Krupovic, M. (2016a) Bipartite network analysis of the archaeal virosphere: evolutionary connections between viruses and capsidless mobile elements. *Journal of Virology*, 90, 11043–11055.

Iranzo, J., Krupovic, M. & Koonin, E.V. (2016b) The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. *mBio*, 7, e00978-16.

Kanehisa, M., Sato, Y. & Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, 428, 726–731.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. et al. (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359.

Kieft, K., Zhou, Z. & Anantharaman, K. (2020) VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*, 8, 90.

Kim, J.-G., Kim, S.-J., Cvirkaite-Krupovic, V., Yu, W.-J., Gwak, J.-H., López-Pérez, M. et al. (2019) Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 15645–15650.

Köster, J. & Rahmann, S. (2012) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28, 2520–2522.

Krupovič, M. & Bamford, D.H. (2008) Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum Euryarchaeota. *Virology*, 375, 292–300.

Krupovič, M., Forterre, P. & Bamford, D.H. (2010) Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *Journal of Molecular Biology*, 397, 144–160.

Krupovič, M., Gribaldo, S., Bamford, D.H. & Forterre, P. (2010a) The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Molecular Biology and Evolution*, 27, 2716–2732.

Krupovic, M., Quemin, E.R.J., Bamford, D.H., Forterre, P. & Prangishvili, D. (2014) Unification of the globally distributed spindle-shaped viruses of the archaea. *Journal of Virology*, 88, 2354–2358.

Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D. & Koonin, E.V. (2018) Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Virus Research*, 244, 181–193.

Krupovic, M., Dolja, V.V. & Koonin, E.V. (2020) The LUCA and its complex virome. *Nature Reviews Microbiology*, 18, 661–670.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.

Lei, Y., Wei, L., Liu, T., Xiao, Y., Dang, Y., Sun, D. et al. (2018) Magnetite enhances anaerobic digestion and methanogenesis of fresh leachate from a municipal solid waste incineration plant. *Chemical Engineering Journal*, 348, 992–999.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Liang, B., Wang, L.-Y., Mbadinga, S.M., Liu, J.-F., Yang, S.-Z., Gu, J.-D. et al. (2015) Anaerolineaceae and Methanosaeta turned to be the dominant microorganisms in alkanes-dependent methanogenic culture after long-term of incubation. *AMB Express*, 5, 37.

Lin, Q., De Vrieze, J., Li, J. & Li, X. (2016) Temperature affects microbial abundance, activity and interactions in anaerobic digestion. *Bioresource Technology*, 209, 228–236.

Liu, Y., Demina, T.A., Roux, S., Aiewsakun, P., Kazlauskas, D., Simmonds, P. et al. (2021) Diversity, taxonomy, and evolution of archaeal viruses of the class Caudoviricetes. *PLoS Biology*, 19, e3001442.

Lopes, A., Tavares, P., Petit, M.-A., Guérois, R. & Zinn-Justin, S. (2014) Automated classification of tailed bacteriophages according to their neck organization. *BMC Genomics*, 15, 1027.

Luo, Y., Pfister, P., Leisinger, T. & Wasserfallen, A. (2001) The genome of archaeal prophage ΨM100 encodes the lytic enzyme responsible for autolysis of Methanothermobacter wolfeii. *Journal of Bacteriology*, 183, 5788–5792.

Lyu, Z., Shao, N., Akinyemi, T. & Whitman, W.B. (2018) Methanogenesis. *Current Biology*, 28, R727–R732.

Madigou, C., Lê Cao, K.-A., Bureau, C., Mazéas, L., Déjean, S. & Chapleur, O. (2019) Ecological consequences of abrupt temperature changes in anaerobic digesters. *Chemical Engineering Journal*, 361, 266–277.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P. et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, 9, 467–477.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O. S., Brouns, S.J.J. et al. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18, 67–83.

Marbouty, M., Cournac, A., Flot, J.-F., Marie-Nelly, H., Mozziconacci, J. & Koszul, R. (2014) Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife*, 3, e03318.

von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F. H. & Dutilh, B.E. (2019) Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology*, 20, 217.

Meile, L., Jenal, U., Studer, D., Jordan, M. & Leisinger, T. (1989) Characterization of ψM1, a virulent phage of Methanobacterium thermoautotrophicum Marburg. *Archives of Microbiology*, 152, 105–110.

Menzel, P., Ng, K.L. & Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7, 11257.

Molnár, J., Magyar, B., Schneider, G., Laczi, K., Valappil, S.K., Kovács, Á.L. et al. (2020) Identification of a novel archaea virus, detected in hydrocarbon polluted Hungarian and Canadian samples. *PLoS One*, 15, e0231864.

Nasko, D.J., Ferrell, B.D., Moore, R.M., Bhavsar, J.D., Polson, S. W. & Wommack, K.E. (2019) CRISPR spacers indicate preferential matching of specific virioplankton genes. *mBio*, 10, e02651-18.

Nayfach, S., Camargo, A.P., Eloe-Fadrosh, E., Roux, S. & Kyrpides, N. (2021) CheckV: assessing the quality of metagenome-assembled viral genomes. *Nature Biotechnology*, 39, 578–585.

Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W. et al. (2007) DNA stable-isotope probing. *Nature Protocols*, 2, 860–866.

Nölling, J., Groffen, A. & de Vos, W.M. (1993) φ F1 and φF3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus Methanobacterium. *Microbiology*, 139, 2511–2516.
Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. (2016) metaSPAdes: a new versatile de novo metagenomics assembler. arXiv:160403071 [q-bio].
Olo Ndela, E., Enault, F. & Toussaint, A. (2021) Transposable prophages in Leptospira: An ancient, now diverse, group predominant in causative agents of Weil's disease. *International Journal of Molecular Sciences*, 22, 13434.
Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.-D. et al. (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium Sulfolobus shibatae. *Virology*, 185, 242–250.
Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. & Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25, 1043–1055.
Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N. et al. (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2, 1533–1542.
Pfister, P., Wasserfallen, A., Stettler, R. & Leisinger, T. (1998) Molecular analysis of Methanobacterium phage ΨM2. *Molecular Microbiology*, 30, 233–244.
Philosof, A., Yutin, N., Flores-Uribe, J., Sharon, I., Koonin, E.V. & Béjà, O. (2017) Novel abundant oceanic viruses of uncultured marine group II Euryarchaeota. *Current Biology*, 27, 1362–1368.
Pickard, D.J.J. (2009) Preparation of bacteriophage lysates and pure DNA. *Methods in Molecular Biology*, 502, 3–9.
Pina, M., Bize, A., Forterre, P. & Prangishvili, D. (2011) The archeoviruses. *FEMS Microbiology Reviews*, 35, 1035–1054.
Poirier, S., Desmond-Le Quéméner, E., Madigou, C., Bouchez, T. & Chapleur, O. (2016) Anaerobic digestion of biowaste under extreme ammonia concentration: identification of key microbial phylotypes. *Bioresource Technology*, 207, 92–101.
Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V. & Krupovic, M. (2017) The enigmatic archaeal virosphere. *Nature Reviews Microbiology*, 15, 724–739.
Puig-Castellví, F., Midoux, C., Guenne, A., Conteau, D., Franchi, O., Bureau, C. et al. (2022) Metataxonomics, metagenomics and metabolomics analysis of the influence of temperature modification in full-scale anaerobic digesters. *Bioresource Technology*, 346, 126612.
Radajewski, S., Ineson, P., Parekh, N.R. & Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature*, 403, 646–649.
Sakowski, E.G., Arora-Williams, K., Tian, F., Zayed, A.A., Zablocki, O., Sullivan, M.B. et al. (2021) Interaction dynamics and virus–host range for estuarine actinophages captured by epicPCR. *Nature Microbiology*, 6, 630–642.
Si, B., Liu, Z., Zhang, Y., Li, J., Shen, R., Zhu, Z. et al. (2016) Towards biohythane production from biomass: influence of operational stage on anaerobic fermentation and microbial community. *International Journal of Hydrogen Energy*, 41, 4429–4438.
Snyder, J.C., Bolduc, B. & Young, M.J. (2015) 40 Years of archaeal virology: expanding viral diversity. *Virology*, 479–480, 369–378.
Spencer, S.J., Tamminen, M.V., Preheim, S.P., Guo, M.T., Briggs, A. W., Brito, I.L. et al. (2016) Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *The ISME Journal*, 10, 427–436.
Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J. & Söding, J. (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*, 20, 1–15.
Sun, H., Yang, Z., Shi, G., Arhin, S.G., Papadakis, V.G., Goula, M.A. et al. (2021) Methane production from acetate, formate and H_2/CO_2 under high ammonia level: modified ADM1 simulation and microbial characterization. *Science of the Total Environment*, 783, 147581.
Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R. E., Mom, R. et al. (2021) PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics*, 3, lqab067.
Thauer, R.K., Kaster, A.-K., Seedorf, H., Buckel, W. & Hedderich, R. (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Reviews Microbiology*, 6, 579–591.
Thiroux, S., Dupont, S., Nesbø, C.L., Bienvenu, N., Krupovic, M., L'Haridon, S. et al. (2021) The first head-tailed virus, MFTV1, infecting hyperthermophilic methanogenic deep-sea archaea. *Environmental Microbiology*, 23, 3614–3626.
Wang, Y. & Qian, P.-Y. (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One*, 4, e7401.
Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261–5267.
Wang, J., Zhao, Z. & Zhang, Y. (2021) Enhancing anaerobic digestion of kitchen wastes with biochar: link between different properties and critical mechanisms of promoting interspecies electron transfer. *Renewable Energy*, 167, 791–799.
Weidenbach, K., Nickel, L., Neve, H., Alkhnbashi, O.S., Künzel, S., Kupczok, A. et al. (2017) Methanosarcina spherical virus, a novel archaeal lytic virus targeting Methanosarcina strains. *Journal of Virology*, 91, e00955-17.
Weidenbach, K., Wolf, S., Kupczok, A., Kern, T., Fischer, M.A., Reetz, J. et al. (2021) Characterization of Blf4, an archaeal lytic virus targeting a member of the Methanomicrobiales. *Viruses*, 13, 1934.
Whiticar, M.J., Faber, E. & Schoell, M. (1986) Biogenic methane formation in marine and freshwater environments: CO_2 reduction vs. acetate fermentation—isotope evidence. *Geochimica et Cosmochimica Acta*, 50, 693–709.
Wolf, S., Fischer, M.A., Kupczok, A., Reetz, J., Kern, T., Schmitz, R. A. et al. (2019) Characterization of the lytic archaeal virus Drs3 infecting Methanobacterium formicicum. *Archives of Virology*, 164, 667–674.
Wood, A.G., Whitman, W.B. & Konisky, J. (1989) Isolation and characterization of an archaebacterial viruslike particle from Methanococcus voltae A3. *Journal of Bacteriology*, 171, 93–98.
Zhang, R., Mirdita, M., Levy Karin, E., Norroy, C., Galiez, C. & Söding, J. (2021) SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. *Bioinformatics*, 37, 3364–3366.
Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M. et al. (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *Journal of Molecular Biology*, 430, 2237–2243.
Zolfo, M., Pinto, F., Asnicar, F., Manghi, P., Tett, A., Bushman, F.D. et al. (2019) Detecting contamination in viromes using ViromeQC. *Nature Biotechnology*, 37, 1408–1412.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ngo, V.Q.H., Enault, F., Midoux, C., Mariadassou, M., Chapleur, O., Mazéas, L. et al. (2022) Diversity of novel archaeal viruses infecting methanogens discovered through coupling of stable isotope probing and metagenomics. *Environmental Microbiology*, 1–16. Available from: https://doi.org/10.1111/1462-2920.16120